

# MMH\*, Home Work 1 HW1

Elwannas HIRI<sup>†</sup>

January 9, 2021

**Homework content** In this homework we will try to answer some essentially question to be able to write an article about *Build a parallel Corpora between Algerian Dialect (AD) and Modern Standard Arabic(MSA) for Machine Translation*<sup>1</sup> and publish it.

## 1 Subject Area:

The field of the project is *Natural Language Processing*(NLP).

## 2 Aim:

Build Corpora for Machine Translation(MT) projects

## 3 Arguments:

Create a Corpora of parrallel AD and MSA for developpers. lack of resources is main arguments and motivation to build the Corpora. for other regions such as Marroco, Tunisia, Lebanon, Middle-East Regions and Egypt, they have their own Corporas for example MADAR, it contains some of AD but not rich.

---

\*Methods Meta-heuristic

<sup>†</sup>Computer science student at Ahmed Draia University -Adrar-, github:  
<https://github.com/t1g3r01>

<sup>1</sup>MMH mini-project subject

## 4 Objectives:

- Collect as much as possible for parrallel MSA-DA dataset of Algerian dialect.
- Filter collected data.
- Organize the dataset depends on location.
- Build a parallel dataset ready for the use for future works.